



日 本 国 特 許 庁
JAPAN PATENT OFFICE

MITSUBI
March 23, 2004
BSK B, CCP
703-205-8000
1163-0500PUS
1041

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日 2 0 0 3 年 4 月 1 0 日
Date of Application:

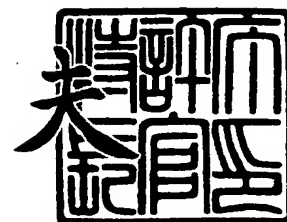
出 願 番 号 特 願 2 0 0 3 - 1 0 6 8 7 8
Application Number:
[ST. 10/C]: [J P 2 0 0 3 - 1 0 6 8 7 8]

出 願 人 三 菱 電 機 株 式 有 限 公 司
Applicant(s):

2 0 0 3 年 1 1 月 2 5 日

特 許 庁 長 官
Commissioner,
Japan Patent Office

今 井 康 夫



出 証 番 号 出 証 特 2 0 0 3 - 3 0 9 7 0 6 3

【書類名】 特許願

【整理番号】 545166JP01

【提出日】 平成15年 4月10日

【あて先】 特許庁長官 殿

【国際特許分類】 G06F 17/27
G06F 17/28

【発明者】

【住所又は居所】 東京都千代田区丸の内二丁目2番3号 三菱電機株式会社
社内

【氏名】 三次 達也

【特許出願人】

【識別番号】 000006013

【氏名又は名称】 三菱電機株式会社

【代理人】

【識別番号】 100066474

【弁理士】

【氏名又は名称】 田澤 博昭

【選任した代理人】

【識別番号】 100088605

【弁理士】

【氏名又は名称】 加藤 公延

【手数料の表示】

【予納台帳番号】 020640

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 文章情報処理装置

【特許請求の範囲】

【請求項 1】 プレーンな文章を入力するプレーン文章入力部と、
形態要素解析及び構文解析に使用する辞書を記憶する辞書記憶部と、
前記プレーン文章入力部から入力されたプレーンな文章を前記辞書記憶部に記憶された辞書を用いて形態要素解析することによりトークンに分解する形態要素解析部と、
前記プレーンな文章の構文に基づき、前記形態要素解析部で得られたトークンの品詞を解析して意味のある語を含む構造化された文章を生成する構文解析部と、
マークアップに使用されるデータを記憶するデータ記憶部と、
前記構文解析部で生成された構造化された文章に含まれる意味のある語に対して、前記データ記憶部に該意味のある語に関係付けて記憶されているデータを読み出して付加するマークアップを行う要素品格処理部と、
前記要素品格処理部で生成されたマークアップ文章を出力するマークアップ文章出力部
とを備えた文章情報処理装置。

【請求項 2】 プレーンなテキスト文章を生成してプレーン文章入力部に送るテキスト文章入力部を更に備えたことを特徴とする請求項 1 記載の文章情報処理装置。

【請求項 3】 外部から通信によりプレーンな文章を入力してプレーン文章入力部に送る外部通信入力部を更に備えたことを特徴とする請求項 1 記載の文章情報処理装置。

【請求項 4】 外部から通信によりマークアップがなされた文章を入力する外部通信入力部と、
前記外部通信入力部からのマークアップがなされた文章を解析することによりマークアップを除去してプレーンな文章を生成し、マークアップ文章入力部に送るマークアップ文章解析部

とを更に備えたことを特徴とする請求項 1 記載の文章情報処理装置。

【請求項 5】 外部から通信により電子メールを入力する外部通信入力部と

、
前記外部通信入力部からの電子メールの構造を解析してプレーンな文章を生成
、マークアップ文章入力部に送る電子メール構造解析部

とを更に備えたことを特徴とする請求項 1 記載の文章情報処理装置。

【請求項 6】 マークアップ文章出力部からのマークアップ文章を通信により外部に送出する外部通信出力部、を更に備えたことを特徴とする請求項 2 から請求項 5 のうちのいずれか 1 項記載の文章情報処理装置。

【請求項 7】 要素品格処理部は、

構文解析部で生成された構造化された文章に含まれる意味のある語に対して、
意味の内容を表す意味要素を関係付ける要素関係解析部を備えたことを特徴とする
請求項 1 記載の文章情報処理装置。

【請求項 8】 固有名を表す固有名要素にマークアップを付加する固有名要素処理部を更に備え、

要素関係解析部は、意味要素が固有名要素である場合に、前記固有名要素処理部を起動することを特徴とする請求項 7 記載の文章情報処理装置。

【請求項 9】 時刻を表す時間要素にマークアップを付加する時間要素処理部を更に備え、

要素関係解析部は、意味要素が時間要素である場合に、前記時間要素処理部を起動することを特徴とする請求項 7 記載の文章情報処理装置。

【請求項 10】 時間要素に対応する日時を算出する時刻処理部を更に備え、

時間要素処理部は、前記時刻処理部からの日時をマークアップとして前記時間要素に付加することを特徴とする請求項 9 記載の文章情報処理装置。

【請求項 11】 位置を表す位置要素にマークアップを付加する位置要素処理部を更に備え、

要素関係解析部は、意味要素が位置要素である場合に、前記位置要素処理部を起動することを特徴とする請求項 7 記載の文章情報処理装置。

【請求項 12】 位置要素に対応する現在位置を推定する現在位置推定部を更に備え、

位置要素処理部は、前記現在位置推定部からの現在位置をマークアップとして前記位置要素に付加することを特徴とする請求項 11 記載の文章情報処理装置。

【請求項 13】 行為を表す行為要素にマークアップを付加する行為要素処理部を更に備え、

要素関係解析部は、意味要素が行為要素である場合に、前記行為要素処理部を起動することを特徴とする請求項 7 記載の文章情報処理装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

この発明は、文章情報処理装置に関し、特に文章の中からの意味のある単語や連語に、その意味又は内容を表す情報を付加する技術に関するものである。

【0002】

【従来の技術】

従来、テキストデータの中で使用されている単独の単語を統計的に処理することにより単独の単語を自動的に分類するものとして、例えばテキストデータにおいて出現する確率が所定値以上の単語クラス列にトークンを付与し、テキストデータの単語・トークン列に含まれる単語とトークンとが混在する集合を、テキストデータの単語・トークン列の生成確率が最大になるように分割し、トークンをテキストデータに存在する連語に置換し、以て単語と連語とをまとめて自動的に分類する技術が知られている（例えば、特許文献 1 参照）。

【0003】

また、膨大な文章情報を要約したり、音声で聞き取りやすい表現に直したり、逆に話し言葉で書かれた文章を読みやすい書き言葉に変換したり、新聞記事等から人名や地名や組織名や日時などの文章の重要な構成要素（固有表現）を抽出したりするシステムにおいて、文字列の制約を含む書き換え規則や固有表現抽出規則を、処理の順序を気にすることなく、宣言的かつ簡潔に指定することを可能にする技術が知られている（例えば、特許文献 2 参照）。この技術で、翻訳装置に

より、ユーザが記述した書き換え規則の集合を確定節文法の規則の集合に変換し、次に、規則統合装置により、確定節文法の規則の集合を高速並行処理の可能な統合された規則に変換する。書き換え実行装置は、統合された規則と変換すべき文章（原文）を受けとり、変換した結果を出力する。

【0004】

【特許文献1】

特開平10-97286号公報

【特許文献2】

特開2001-67355号公報

【0005】

【発明が解決しようとする課題】

しかしながら、上述した特許文献1や特許文献2に開示された技術では、単語と連語とを自動的に分類するに止まり、単語や連語の持つ意味や内容までを表現することはできない。

【0006】

また、近年の通信技術や情報制御技術の発達に伴い、異なる場所及び時間で、各々の情報機器端末において、電子メールの交換やホームページの閲覧に代表されるような文字情報の交換が頻繁に行われている。しかし、文字情報に含まれる単語及び連語の解読は人間の記憶や判断に基づいて行われており、時として文脈や構文の判断ミス、記憶のミスに起因して文字情報の提供者と文字情報の入手者との間で、文章の意味や内容が異なって理解される場合がある。その結果、文字情報を用いた意志伝達がスムーズに行われないという課題がある。

【0007】

この発明は、上記のような課題を解決するためになされたものであり、文章の意味及び内容を確実に伝えることのできる文章情報処理装置を提供することを目的とする。

【0008】

【課題を解決するための手段】

この発明に係る文章情報処理装置は、プレーンな文章を入力するプレーン文章

入力部と、形態要素解析及び構文解析に使用する辞書を記憶する辞書記憶部と、プレーン文章入力部から入力されたプレーンな文章を辞書記憶部に記憶された辞書を用いて形態要素解析することによりトークンに分解する形態要素解析部と、プレーンな文章の構文に基づき、形態要素解析部で得られたトークンの品詞を解析して意味のある語を含む構造化された文章を生成する構文解析部と、マークアップに使用されるデータを記憶するデータ記憶部と、構文解析部で生成された構造化された文章に含まれる意味のある語に対して、データ記憶部に該意味のある語に関係付けて記憶されているデータを読み出して付加するマークアップを行う要素品格処理部と、要素品格処理部で生成されたマークアップ文章を出力するマークアップ文章出力部とを備えているものである。

【0009】

【発明の実施の形態】

以下、この発明の実施の形態を図面を参照しながら詳細に説明する。なお、この文章情報処理装置は、例えばパーソナルコンピュータ、サーバコンピュータ、PDAといった種々の情報処理装置の上に、ソフトウェア処理によって構築することができる。

実施の形態1.

図1は、この発明の実施の形態1に係る文章情報処理装置の構成を示すブロック図である。この文章情報処理装置は、入力部1、マークアップ処理部2及び出力部3から構成されている。

【0010】

入力部1は、文章情報処理装置が構築される情報処理装置の入力装置から構成することができる。この入力部1は、種々の形態の文章からプレーンな文章を生成してマークアップ処理部2に送る。マークアップ処理部2は、詳細は後述するが、プレーンな文章にマークアップを付したマークアップ文章を生成し、出力部3に送る。出力部3は、マークアップ処理部2から送られてくるマークアップ文章を出力する。

【0011】

マークアップ処理部2は、プレーン文章入力部10、辞書記憶部11、形態要

素解析部 12、構文解析部 13、データ記憶部 14、要素品格処理部 15、マークアップ文章記憶部 16 及びマークアップ文章出力部 17 から構成されている。

【0012】

プレーン文章入力部 10 は、入力部 1 から送られてくるプレーンな文章を受け取って形態要素解析部 12 に送る。辞書記憶部 11 は、形態要素解析部 12 における形態要素解析及び構文解析部 13 における構文解析にそれぞれ使用される辞書を記憶する。

【0013】

形態要素解析部 12 は、プレーン文章入力部 10 から受け取ったプレーンな文章を、辞書記憶部 11 に格納されている辞書を参照しながら形態素解析を行ってトークンに分解する。ここで、トークンとは、最小の意味単位を構成する文字列をいう。この形態要素解析部 12 でトークンに分解されたデータは、構文解析部 13 に送られる。なお、この形態要素解析部 12 で行われるトークン分解の技術は周知であるので、この明細書では詳細な説明は省略する。

【0014】

構文解析部 13 は、形態要素解析部 12 からのトークン分解されたデータを入力し、辞書記憶部 11 に格納されている辞書を参照しながら文章の構文に基づいて品詞を解析する構文解析を実行して文章構造を表すデータを生成する。この構造化された文章には、位置要素、時間要素、固有名詞要素、行為要素といった意味のある語（単語及び連語を含む。以下同じ）が含まれる。この構文解析部 13 で生成された構造化された文章は、要素品格処理部 15 に送られる。なお、この構文解析部 13 で行われる構造化された文章を生成する技術は周知であるので、この明細書では詳細な説明は省略する。

【0015】

データ記憶部 14 は、マークアップに使用される種々のデータをデータベースとして格納する。このデータ記憶部 14 に形成されるデータベースは、入力部 1 から入力されるデータ及びマークアップ処理部 2 の結果によるデータに基づいて随時更新される。

【0016】

要素品格処理部 15 は、構文解析部 13 から送られてくる構造化された文章に含まれる意味のある語に対し、その語の意味又は内容を表すデータを付加するマークアップを実行し、構造化されたマークアップ文章を生成する。この要素品格処理部 15 で生成されたマークアップ文章は、マークアップ文章記憶部 16 に送られる。

【0017】

マークアップ文章記憶部 16 は、要素品格処理部 15 から送られてくるマークアップ文章を記憶する。このマークアップ文章記憶部 16 に格納されたマークアップ文章は、マークアップ文章出力部 17 から読み出される。マークアップ文章出力部 17 は、マークアップ文章記憶部 16 に記憶されているマークアップ文章を読み出し、出力部 3 に送る。

【0018】

ここで、以上のように構成される文章情報処理装置の概要を理解するために、プレーンな文章からマークアップ文章が作成される過程を、製品発表に使用されるプレリリースの文章を例にとって説明する。プレリリース原稿の一部は、以下にクォーテーションで囲って示す文章から成るものとする。

【0019】

“「〇〇株式会社は、カーナビ業界初で J a v a（登録商標）TM アプリに対応した DVD カーナビゲーションシステム「A シリーズ」2 モデルと、高性能ながら普及価格を実現した DVD カーナビゲーションシステム「B シリーズ」2 モデルの合計 4 モデルを 4 月 20 日より発売しますので、お知らせします。」”

【0020】

このプレリリース原稿が入力部 1 からマークアップ処理部 2 に入力されると、プレーン文章入力部 10 を経由して形態要素解析部 12 に供給される。形態要素解析部 12 は、辞書記憶部 11 に記憶された辞書を参照しながら形態要素解析を実行し、以下にクォーテーションで囲って示すトークン分解されたデータを出力する。

【0021】

“/「/〇〇/株式/会社/は、/カーナビ/業界/初/で/J a v a（登録商標）TM

/アプリ/に/対応した/DVD/カー/ナビゲーション/システム/「/A/シリーズ/」/2/モデル/と、/高性能/ながら/普及/価格/を/実現/した/DVD/カー/ナビゲーション/システム/「/B/シリーズ/」/2/モデル/の/合計/4/モデル/を/4/月/20/日/より/発売/します/ので、/お知らせ/します。/」/”

【0022】

この形態要素解析部12から出力されるトークン分解されたデータは、構文解析部13に送られる。構文解析部13は、形態要素解析部12からのデータを受け取り、辞書記憶部11に記憶された辞書を参照しながら構文解析を実行し、下記表1に示す構造化された文章を出力する。

【0023】

[表1]

<文章構造/>

<主語/>〇〇株式会社</主語>

<接続語/>は、</接続語>

<補語/>

<主語/>「Aシリーズ」</主語>

<修飾語/>カーナビ業界初

<接続語/>で</接続語>

<修飾語/>J a v a（登録商標）TMアプリ</修飾語>

<接続語/>に対応した</接続語>

</修飾語>

<修飾語/>DVDカーナビゲーションシステム</修飾語>

<修飾語/>2モデル</修飾語>

<接続語/>と、</接続語>

<主語/>「Bシリーズ」</主語>

<修飾語/>高性能

<接続語/>ながら</接続語>

</修飾語>

<修飾語/>普及価格</修飾語>

<接続語/>を実現した</接続語>

<修飾語/>DVDカーナビゲーションシステム</修飾語>

<修飾語/>2モデル</修飾語>

<接続語/>の</接続語>

<主語/>合計4モデル</主語>

<接続語/>を</接続語>

<修飾語/>4月20日より</修飾語>

<述語/>発売します</述語>

<接続語/>ので</接続語>

</補語>

<述語/>お知らせします。</述語>

</文章構造>

【0024】

この構文解析部13で生成された構造化された文章は、要素品格処理部15に送られる。要素品格処理部15は、構文解析部13からの構造化された文章を入力し、データ記憶部14に形成されたデータベースを参照しながら、マークアップを付加する処理を実行し、下記表2に示すような意味のある語にマークアップがなされたマークアップ文章を生成する。この要素品格処理部15で生成されたマークアップ文章は、マークアップ文章記憶部16に格納される。

【0025】

[表2]

<文章/>

<主語/>〇〇株式会社

<Position/>

<緯度/>+35.676666</緯度>

<経度/>+139.766666</緯度>

</Position>

<Peculiar Name/>

<名称/>〇〇株式会社

<ふりがな/>まるまるかぶしきがいしゃ</ふりがな>

<Name/>MARUMARU CORPORATION</Name>

</名称>

<役員/>

<URI/>

<http://www.marumaru.co.jp/keireki/pdf/2001/p6.pdf>

</URI>

</役員>

<組織/>

<URI/>

<http://www.marumaru.co.jp/keireki/pdf/2001/p4.pdf>

</URI>

</組織>

<住所/>

<郵便番号/>1 0 0 - 1 2 3 4</郵便番号>

<広域地方自治体/>東京都</広域地方自治体>

<地方自治体/>千代田区</地方自治体>

<町名/>丸の内二丁目</町名>

<住居表示/>2 - 3</住居表示>

</住所>

<連絡先/>

<URI/><http://www.marumaru.co.jp></URI>

<TEL/>03-3218-****</TEL>

<FAX/>03-3218-****</FAX>

</連絡先>

</Peculiar Name>

</主語>

<接続語/>は、</接続語>

<補語/>

<主語/>「A シリーズ」
 <Peculiar Name/>
 <URI/>http://www.marumaru.co.jp/carele/carnavi/</URI>
 </Peculiar Name>
</主語>
<修飾語/>カーナビ業界
 <Peculiar Name/><Peculiar Name/>
 <修飾語/>初
 <Times/> 1 </Times>
 </修飾語>
 <接続語/>で</接続語>
 <修飾語/>J a v a （登録商標） T M アプリ
 <Peculiar Name/>
 <URL/>http://www.sun.co.jp/software/java/</URL>
 </Peculiar Name>
 </修飾語>
 <接続語/>に対応した</接続語>
</修飾語>
<修飾語/>D V D カーナビゲーションシステム
 <Peculiar Name/></Peculiar Name>
</修飾語>
<修飾語/>2 モデル</修飾語>
<接続語/>と、</接続語>
<主語/>「B シリーズ」
 <Peculiar Name/>
 <URI/>http://www.marumaru.co.jp/carele/carnavi/</URI>
 </Peculiar Name>
</主語>
<修飾語/>高性能

<接続語/>ながら</接続語>
</修飾語>
<修飾語/>普及価格</修飾語>
<接続語/>を実現した</接続語>
<修飾語/>D V Dカーナビゲーションシステム
 <Peculiar Name/></Peculiar Name>
</修飾語>
<修飾語/>2 モデル</修飾語>
<接続語/>の</接続語>
<主語/>合計 4 モデル</主語>
<接続語/>を</接続語>
<修飾語/>4 月 2 0 日より
 <Time/>
 <From/>
 <DATE/>2002-04-20</DATE>
 </From>
 </Time>
</修飾語>
<述語/>発売します
 <Act/></Act>
</述語>
<接続語/>ので</接続語>
</補語>
<述語/>お知らせします。
 <Act/></Act>
</述語>
</文章>

【 0 0 2 6 】

このように要素品格処理部 1 5 でマークアップがなされることにより、文章の

意味のある語に対しては、その意味又は内容がマークアップされたマークアップ文章が生成される。意味のある単語又は連語のマークアップとは、上述した表1及び表2に示した例では、「〇〇株式会社」に対して「位置」、「名称」、「役員」、「組織」、「住所」「連絡先」等の内容が付加され、各々の内容について説明文の付加又はメタ定義がなされることである。なお、マークアップの対象が別のオブジェクトである場合は、各々の要素をメタ定義のURI (Uniform Resource Identifier) として定義することができる。

【0027】

上述した入力部1は、図2に示すように、上述したプレリリース原稿のようなテキストから成るプレーンな文章を入力するテキスト文章入力部1aから構成することができる。テキスト文章入力部1aは、例えばキーボード、外部記憶媒体といったプレーンなテキスト文章を生成できる種々の機器から構成できる。また、出力部3は、例えばディスプレイ装置やプリンタ装置といった文章を可視化できる種々の装置から構成できる。この場合、出力部3は、図3～図5に示すような、マークアップ文章を外部に送信する例えば通信機器から成る外部通信出力部3aで構成することもできる。

【0028】

また、入力部1は、図3に示すように、上述したプレリリース原稿のようなプレーンな文章を外部の通信機器から入力する外部通信入力部1bから構成することができる。また、出力部3は、マークアップ文章を外部に送信する例えば通信機器から成る外部通信出力部3aで構成できる。

【0029】

また、入力部1は、図4に示すように、例えばインターネット上で使用されるHTML (Hyper Text Markup Language) のようなマークアップがなされた文章を入力する外部通信入力部1cと、この外部通信入力部1cからの文章を解析してマークアップを除去し、プレーンな文章を生成するマークアップ文章解析部1dから構成できる。また、出力部3は、マークアップ文章を外部に送信する例えば通信機器から成る外部通信出力部3aで構成できる。

【0030】

更に、入力部 1 は、図 5 に示すように、電子メールのような構造化された文章を入力する外部通信入力部 1 e と、この外部通信入力部 1 e から入力された文章の構造を解析し、ヘッダ部とボディ部との関連性を保ちつつプレーンな文章に変換する電子メール構造解析部 1 f から構成できる。また、出力部 3 は、マークアップ文章を外部に送信する例えば通信機器から成る外部通信出力部 3 a で構成できる。

【0031】

電子メールのヘッダ部には、プロトコルタイプ (X-STMP-Result)、送信日付 (Date)、発信者 (From)、宛先 (To、CC、BCC)、タイトル (Subject)、メール ID (Message-Id)、MIME 種別 (MIME-Version:1.0)、メールソフトの種別 (X-Mailer)、優先度 (X-Priority)、メッセージの優先度 (X-MSMail-Priority)、ボディのタイプ (Content-Type)、ボディの書式 (Content-Transfer-Encoding) 等を表すデータが含まれる。

【0032】

ボディ部は、メッセージの本体であり、プレーンな文章から構成されている。電子メール構造解析部 1 f は、ヘッダ部に含まれるボディのタイプ (Content-Type) を参照してボディ部がテキストであるかどうかを判断し、テキストである場合はプレーンな文章としてそのまま出力し、テキストでない場合はマークアップを削除してプレーンな文章に変換して出力する。また、電子メール構造解析部 1 f は、送信日付 (Date)、発信者 (From)、宛先 (To、CC、BCC)、タイトル (Subject) といった情報をデータ記憶部 1 4 に形成されたデータベースに格納する。

【0033】

今、友人 (山田太郎) が、私 (川上花子) に、2002 年 12 月 25 日 20 時 00 分に、「明日の午後 7 時に、あなたの家に行く。」というメッセージを送る場合を考える。この場合、ヘッダ部には、発信者として山田太郎の電子メールアドレスが、宛先に川上花子の電子メールアドレスが、送信日付に 2002 年 12 月 20 時 00 分がそれぞれ記載される。

【0034】

この電子メールが、図5に示した外部通信入力部1eで受信されて電子メール構造解析部1fでプレーンな文章に変化された後、プレーン文章入力部10を経由して形態要素解析部12に入力されると、形態要素解析部12は、辞書記憶部11に記憶された辞書を参照しながら形態要素解析を行い、以下にクォーテーションで囲って示すトークン分解されたデータを出力する。

【0035】

“/「/明日/の/午後/7/時/に/、/あなた/の/家/に/行く/。/」/”

【0036】

この形態要素解析部12から出力されるトークン分解されたデータは、構文解析部13に送られる。構文解析部13は、形態要素解析部12からのデータを受け取り、辞書記憶部11に記憶された辞書を参照しながら構文解析を実行し、下記表3に示す構造化された文章を出力する。

【0037】

[表3]

<文章構造/>

<主語/></主語>

<補語/>

<修飾語/>明日の午後7時に</修飾語>

<修飾語/>あなたの家に</修飾語>

</補語>

<述語/>行く。</述語>

</文章構造>

【0038】

この構文解析部13で生成された構造化された文章は、要素品格処理部15に送られる。要素品格処理部15は、構文解析部13からの構造化された文章を入力し、データ記憶部14に記憶されたデータを参照しながら、マークアップを付加する処理を実行し、下記表4に示すような意味のある語にマークアップがなされたマークアップ文章を生成する。この要素品格処理部15で生成されたマークアップ文章は、マークアップ文章記憶部16に格納される。

【0039】

[表 4]

<文章/>

<主語/>

<Peculiar Name/>

<名称/>山田太郎

<ふりがな/>やまだたろう</ふりがな>

<Name/>YAMADA TARO</Name>

</名称>

<住所/>

<郵便番号/>100-1234</郵便番号>

<広域地方自治体/>東京都</広域地方自治体>

<地方自治体/>千代田区</地方自治体>

<町名/>丸の内二丁目</町名>

<住居表示/>2-3</住居表示>

</住所>

<連絡先/>

<URI/>yamada taro@malco.com</URI>

<TEL/>03-3218-****</TEL>

<FAX/>03-3218-****</FAX>

</連絡先>

</Peculiar Name>

</主語>

<補語/>

<修飾語/>明日の午後7時に

<Time/>

<At/>

<DATE/>2002-12-26T19:00Z</DATE>

</At>

</Time>
</修飾語>
<修飾語/>あなたの家に
<Peculiar Name/>
 <名称/>川上花子
 <ふりがな/>かわかみはなこ</ふりがな>
 <Name/>KAWAKAMI HANAKO</Name>
 </名称>
 <住所/>
 <郵便番号/>6 6 9 - 5 6 7 8</郵便番号>
 <広域地方自治体/>兵庫県</広域地方自治体>
 <地方自治体/>三田市</地方自治体>
 <町名/>三輪二丁目</町名>
 <住居表示/>3 - 3 3</住居表示>
 </住所>
 <連絡先/>
 <URI/>kawano hanako@malco.com</URI>
 <TEL/>079-559-****</TEL>
 <FAX/>079-559-****</FAX>
 </連絡先>
</Peculiar Name>
</修飾語>
</補語>
<述語/>行く。
 <Act/></Act>
</述語>
</文章>

【 0 0 4 0 】

このように電子メールのメッセージは、意味のある語に対しては、その意味又

は内容がマークアップされる。意味のある語のマークアップとは、上述した例では、メッセージ中では明記されていなかった主語として、「名称」、住所」、「連絡先」といったデータが付加され、各々のデータについての説明文又はメタ定義がなされる。なお、マークアップの対象が別のオブジェクトである場合は、各々の要素をメタ定義のURIとして定義することができる。

【0041】

次に、上述したマークアップ処理部2の一部を構成する要素品格処理部15について、詳細に説明する。

【0042】

要素品格処理部15は、図6に示すように、要素関係解析部20、固有名要素処理部21、時間要素処理部22、位置要素処理部23、行為要素処理部24及びデータ収集部25から構成されている。

【0043】

データ収集部25には、現在位置推定部30、時刻処理部31及び上述したデータ記憶部14が接続されている。データ収集部25は、これら現在位置推定部30、時刻処理部31及びデータ記憶部14からデータを取得して、固有名要素処理部21、時間要素処理部22、位置要素処理部23及び行為要素処理部24における処理に提供する。

【0044】

現在位置推定部30は、文章作成者及び文章に表れる位置要素の現在位置を推定し、要素品格処理部15のデータ収集部25に送る。また、時刻処理部31は、現在の日時に基づき時刻要素（後述する）に対応する日時を算出し、要素品格処理部15のデータ収集部25に送る。

【0045】

要素品格処理部15は、構文解析部13から送られてくる構文解析済み文章に含まれる語の各々について、図7のフローチャートに示す手順に従った処理を実行する。即ち、要素品格処理部15では、先ず、構文解析部13から送られてくる構文解析済み文章に含まれる語が名詞要素であるかどうか調べられる（ステップST10）。そして、名詞要素であることが判断されると、要素関係解析部

20が起動され、要素関係解析処理が実行される（ステップST11）。

【0046】

一方、上記ステップST10で、名詞要素でないことが判断されると、次いで、上記語が動詞要素であるかどうか調べられる（ステップST12）。そして、動詞要素であることが判断されると、要素関係解析部20が起動され、要素関係解析処理が実行される（ステップST11）。一方、上記語が動詞要素でもないことが判断されると、その語は意味要素でないことが認識され、当該語に対する処理を終了して次の語の処理に移る。以下、構文解析部13から送られてくる構文解析済み文章に含まれる全ての語についての処理が終了するまで、ステップST10～ST12の処理が繰り返される。

【0047】

要素関係解析部20は、上述した要素品格処理部15において、構文解析済み文章に含まれる語が名詞要素又は動詞要素であると判断された場合に、更に位置要素、時間要素、固有名要素及び動詞要素といった意味要素の何れかであるかを判断して、各意味要素に対応する処理を実行する。

【0048】

具体的には、要素関係解析部20は、図8のフローチャートに示す手順に従った処理を実行する。即ち、要素関係解析部20では、先ず、構文解析済み文章に含まれる語が名詞要素であるかどうか調べられる（ステップST20）。そして、名詞要素であることが判断されると、次いで、その名詞要素は、位置要素であるかどうか調べられる（ステップST21）。位置名要素には、文章作成者の位置及び文章に表れる位置が含まれる。ここで、位置要素であることが判断されると、位置要素処理部23に制御が移され、位置要素処理が実行される（ステップST22）。その後、シーケンスは上述した要素品格処理部15の処理にリターンする。

【0049】

上記ステップST21で、位置要素でないことが判断されると、次いで、名詞要素は、時間要素であるかどうか調べられる（ステップST23）。時間名要素には、文章作成の時間及び文章に表れる時間が含まれる。ここで、時間要素で

あることが判断されると、時間要素処理部 22 に制御が移され、時間要素処理が実行される（ステップ S T 24）。その後、シーケンスは上述した要素品格処理部 15 の処理にリターンする。

【0050】

上記ステップ S T 23 で、時間要素でないことが判断されると、次いで、名詞要素は、固有名要素であるかどうか調べられる（ステップ S T 25）。固有名要素には、文章作成者の固有名及び文章に表れる固有名が含まれる。ここで、固有名要素であることが判断されると、固有名要素処理部 21 に制御が移され、固有名要素処理が実行される（ステップ S T 26）。その後、シーケンスは上述した要素品格処理部 15 の処理にリターンする。ステップ S T 25 で、固有名要素でないことが判断された場合もシーケンスは上述した要素品格処理部 15 の処理にリターンする。

【0051】

上記ステップ S T 20 で、構文解析済み文章に含まれる語が名詞要素でないことが判断されると、次いで、その語が動詞要素であるかどうか調べられる（ステップ S T 27）。そして、動詞要素であることが判断されると、行為要素処理部 24 に制御が移され、行為要素処理が実行される（ステップ S T 28）。その後、シーケンスは上述した要素品格処理部 15 の処理にリターンする。ステップ S T 27 で、動詞要素でないことが判断された場合もシーケンスは上述した要素品格処理部 15 の処理にリターンする。

【0052】

固有名要素処理部 21 は、固有名要素に対するマークアップ処理を行う。具体的には、固有名要素処理部 21 は、図 9 のフローチャートに示す手順に従った処理を実行する。即ち、固有名要素処理部 21 では、先ず、固有名要素のマークアップ要素データが収集される（ステップ S T 30）。具体的には、固有名要素処理部 21 は、その固有名要素に関係付けられたデータを、データ記憶部 14 に形成されたデータベースからデータ収集部 25 を介して収集する。

【0053】

次いで、固有名要素マークアップの付与が行われる（ステップ S T 31）。即

ち、固有名要素処理部 21 は、上記ステップ S T 30 で収集したデータを固有名要素に対応付ける。次いで、内部データと該テキストとの関係付けマークアップが行われる（ステップ S T 32）。具体的には、固有名要素処理部 21 は、収集したデータを文章構造内にマークアップとして収容する。これにより、例えば表 2 に示すように、「〇〇株式会社」のマークアップとして、タグ Peculiar Name に囲まれた「位置 (Position)」、「名称」、「役員」、「組織」、「住所」、「連絡先」等がタグを付されて組み込まれる。その後、シーケンスは、要素関係解析部 20 の処理にリターンする。

【0054】

位置要素処理部 23 は、位置要素に対するマークアップ処理を行う。具体的には、位置要素処理部 23 は、図 10 のフローチャートに示す手順に従った処理を実行する。即ち、位置要素処理部 23 では、先ず、位置要素のマークアップ要素データが収集される（ステップ S T 40）。具体的には、位置要素処理部 23 は、その位置要素に対応する現在位置を現在位置推定部 30 からデータ収集部 25 を介して収集し、更に、この収集した現在位置に関係付けられたデータをデータ記憶部 14 に形成されたデータベースからデータ収集部 25 を介して収集する。

【0055】

次いで、位置要素マークアップの付与が行われる（ステップ S T 41）。即ち、位置要素処理部 23 は、上記ステップ S T 40 で収集したデータを位置要素に対応付ける。次いで、内部データと該テキストの関係付けマークアップが行われる（ステップ S T 42）。具体的には、位置要素処理部 23 は、収集したデータをテキスト文章構造内にマークアップとして収容する。これにより、例えば表 2 に示すように、「〇〇株式会社」のマークアップとして、タグ Position に囲まれた「経度」、「緯度」といったデータがタグを付されて組み込まれる。その後、シーケンスは、要素関係解析部 20 の処理にリターンする。

【0056】

時間要素処理部 22 は、時間要素に対するマークアップ処理を行う。具体的には、時間要素処理部 22 は、図 11 のフローチャートに示す手順に従った処理を実行する。即ち、時間要素処理部 22 では、先ず、時間要素のマークアップ要素

データの収集が行われる（ステップ S T 5 0）。具体的には、時間要素処理部 22 は、その時間要素に対応する日時を時刻処理部 31 からデータ収集部 25 を介して収集する。

【0057】

次いで、時間要素マークアップの付与が行われる（ステップ S T 5 1）。即ち、時間要素処理部 22 は、表 4 に示した例では、上記ステップ S T 5 0 で収集した日時を時間要素に対応付ける。次いで、内部データと該テキストの関係付けマークアップが行われる（ステップ S T 5 2）。具体的には、時間要素処理部 22 は、収集した日時をテキスト文章構造内にマークアップとして収容する。これにより、例えば表 4 に示すように、「明日の午後 7 時に」のマークアップとして、タグ Time で囲まれた「2002-12-26T19:00Z」といったデータがタグを付されて組み込まれる。その後、シーケンスは、要素関係解析部 20 の処理にリターンする。

【0058】

行為要素処理部 24 は、行為要素に対するマークアップ処理を行う。具体的には、行為要素処理部 24 は、図 12 のフローチャートに示す手順に従った処理を実行する。即ち、行為要素処理部 24 では、先ず、行為要素のマークアップ要素データの収集が行われる（ステップ S T 6 0）。具体的には、行為要素処理部 24 は、データ収集部 25 を介して、その行為に関係付けられたデータをデータ記憶部 14 に形成されたデータベースから収集する。

【0059】

次いで、行為要素マークアップの付与が行われる（ステップ S T 6 1）。即ち、行為要素処理部 24 は、上記ステップ S T 6 0 で収集したデータを行為要素に対応付ける。次いで、内部データと該テキストの関係付けマークアップが行われる（ステップ S T 6 2）。具体的には、行為要素処理部 24 は、収集したデータをテキスト文章構造内にマークアップとして収容する。これにより、例えば表 4 に示すように、「行く」のマークアップとして、タグ Act に囲まれた空要素が組み込まれる。その後、シーケンスは、要素関係解析部 20 の処理にリターンする。

【0060】

以上説明したように、この実施の形態1に係る文章情報処理装置によれば、様々な文字情報や文章を語（単語又は連語）に分解して品詞により分類し、文章構造や情報機器端末に蓄積されているデータベースに基づいて、分類された語に対して意味付け又は内容付けのためのマークアップを施すことにより、文章に含まれている語の意味又は内容が自動的に付加されるので、文章の受け手に対して単語や連語の意味の曖昧さをなくし、意味内容の正確な伝送を行うことができる。その結果、文字情報を用いた意志伝達をスムーズに行うことができる。

【0061】

なお、マークアップされた単語や連語の意味又は内容は文章表現だけでなく、音声や映像などのオブジェクトをも表現できる。従って、様々な文字情報、文章の交換時において、意志伝達をスムーズに行うことができる。

【0062】

なお、上述した実施の形態1では、文章情報処理装置は情報処理装置の上に構築されるものとして説明したが、独立したハードウェアで構成するは勿論可能である。

【0063】

【発明の効果】

以上のように、この発明によれば、文字情報や文章を語に分解して品詞により分類し、文章構造や蓄積されているデータに基づいて、分類された語に対して意味付け又は内容付けのためのマークアップを施すことにより、文章に含まれている語の意味又は内容が自動的に付加されるので、文章の受け手に対して語の意味の曖昧さをなくし、部の意味及び内容を正確に伝えることができる効果がある。

【図面の簡単な説明】

【図1】 この発明の実施の形態1に係る文章情報処理装置の構成を示すブロック図である。

【図2】 この発明の実施の形態1に係る文章情報処理装置の変形例の構成を示すブロック図である。

【図3】 この発明の実施の形態1に係る文章情報処理装置の他の変形例の

構成を示すブロック図である。

【図 4】 この発明の実施の形態 1 に係る文章情報処理装置の更に他の変形例の構成を示すブロック図である。

【図 5】 この発明の実施の形態 1 に係る文章情報処理装置の更に他の変形例の構成を示すブロック図である。

【図 6】 図 1 に示した要素品格処理部の構成を示すブロック図である。

【図 7】 図 6 に示した要素品格処理部で実行される処理を示すフローチャートである。

【図 8】 図 6 に示した要素関係解析部で実行される処理を示すフローチャートである。

【図 9】 図 6 に示した固有名要素処理部で実行される処理を示すフローチャートである。

【図 10】 図 6 に示した位置要素処理部で実行される処理を示すフローチャートである。

【図 11】 図 6 に示した時間要素処理部で実行される処理を示すフローチャートである。

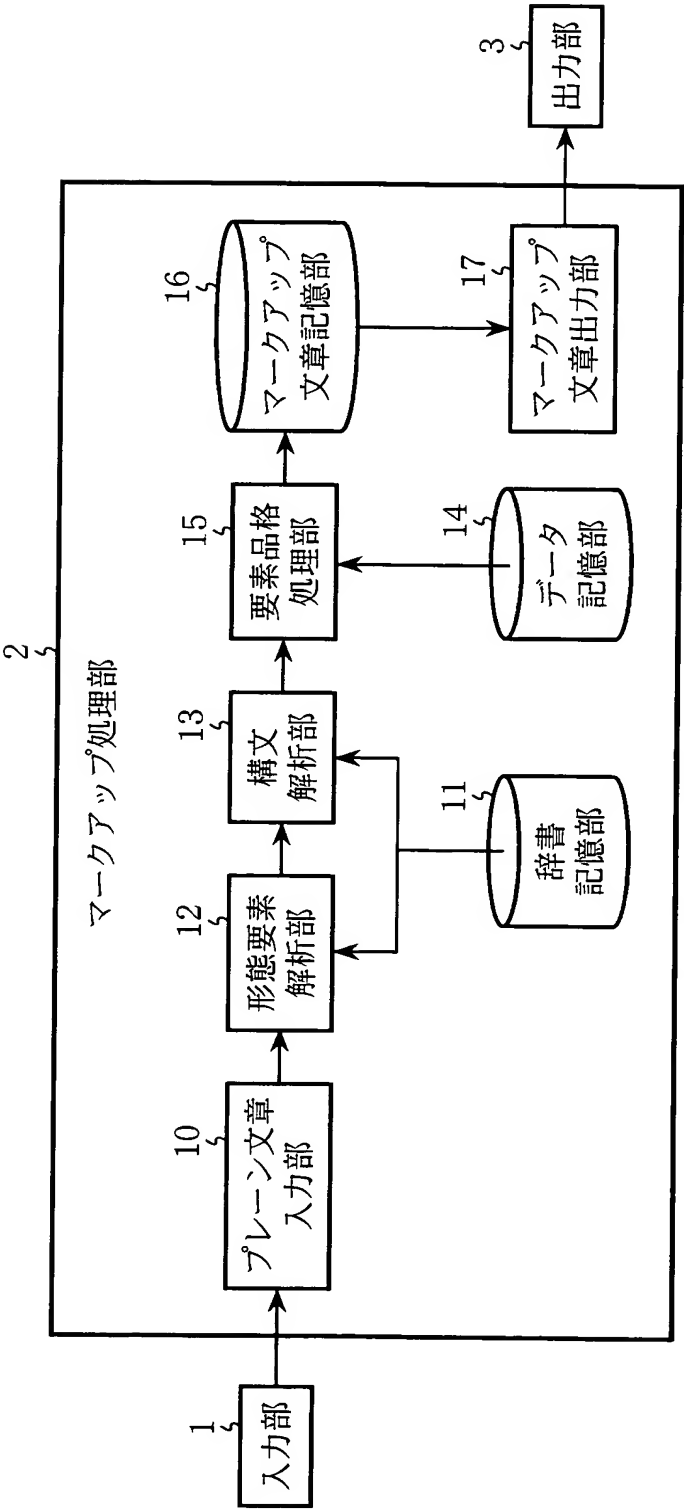
【図 12】 図 6 に示した行為要素処理部で実行される処理を示すフローチャートである。

【符号の説明】

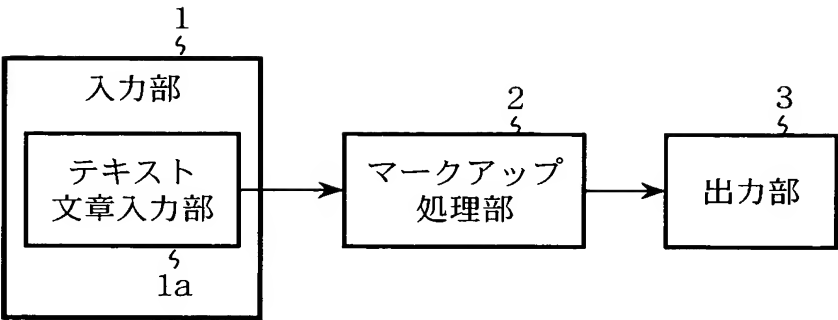
1 入力部、1 a テキスト文章入力部、1 b, 1 c, 1 e 外部通信入力部、1 d マークアップ文書解析部、1 f 電子メール構造解析部、2 マークアップ処理部、3 出力部、3 a 外部通信出力部、10 プレーン文章入力部、11 辞書記憶部、12 形態要素解析部、13 構文解析部、14 データ記憶部、15 要素品格処理部、16 マークアップ文章記憶部、17 マークアップ文章出力部、20 要素関係解析部、21 固有名要素処理部、22 時間要素処理部、23 位置要素処理部、24 行為要素処理部、25 データ収集部、30 現在位置推定部、31 時刻処理部。

【書類名】 図面

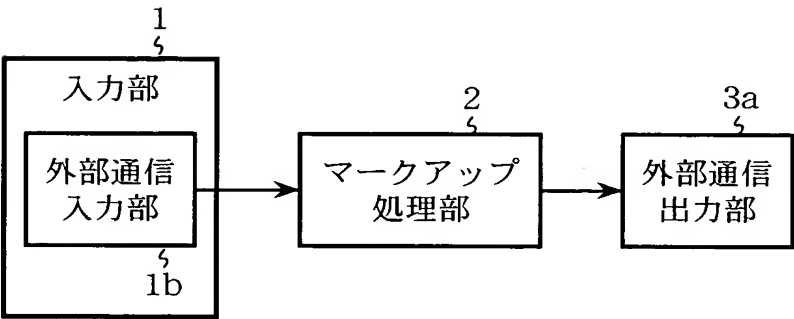
【図 1】



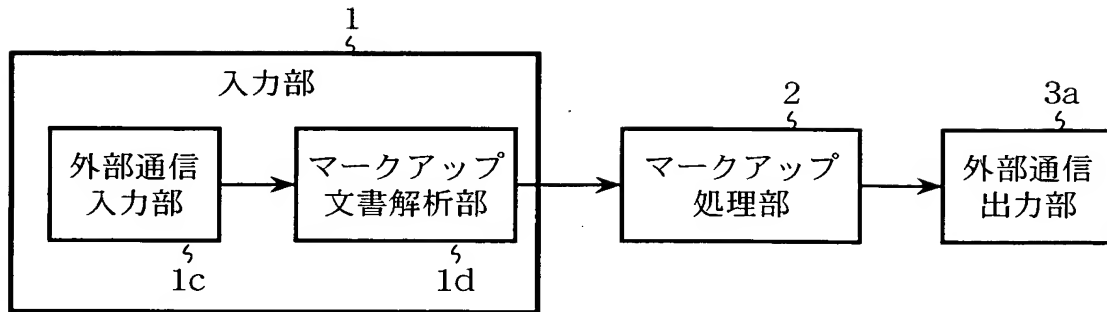
【図 2】



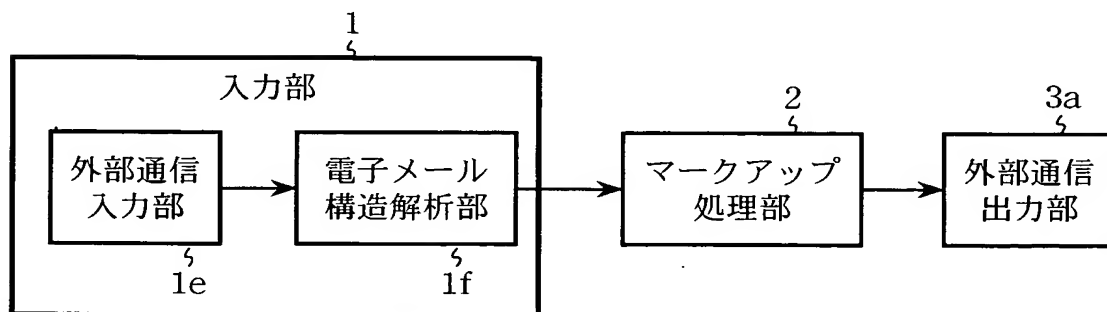
【図 3】



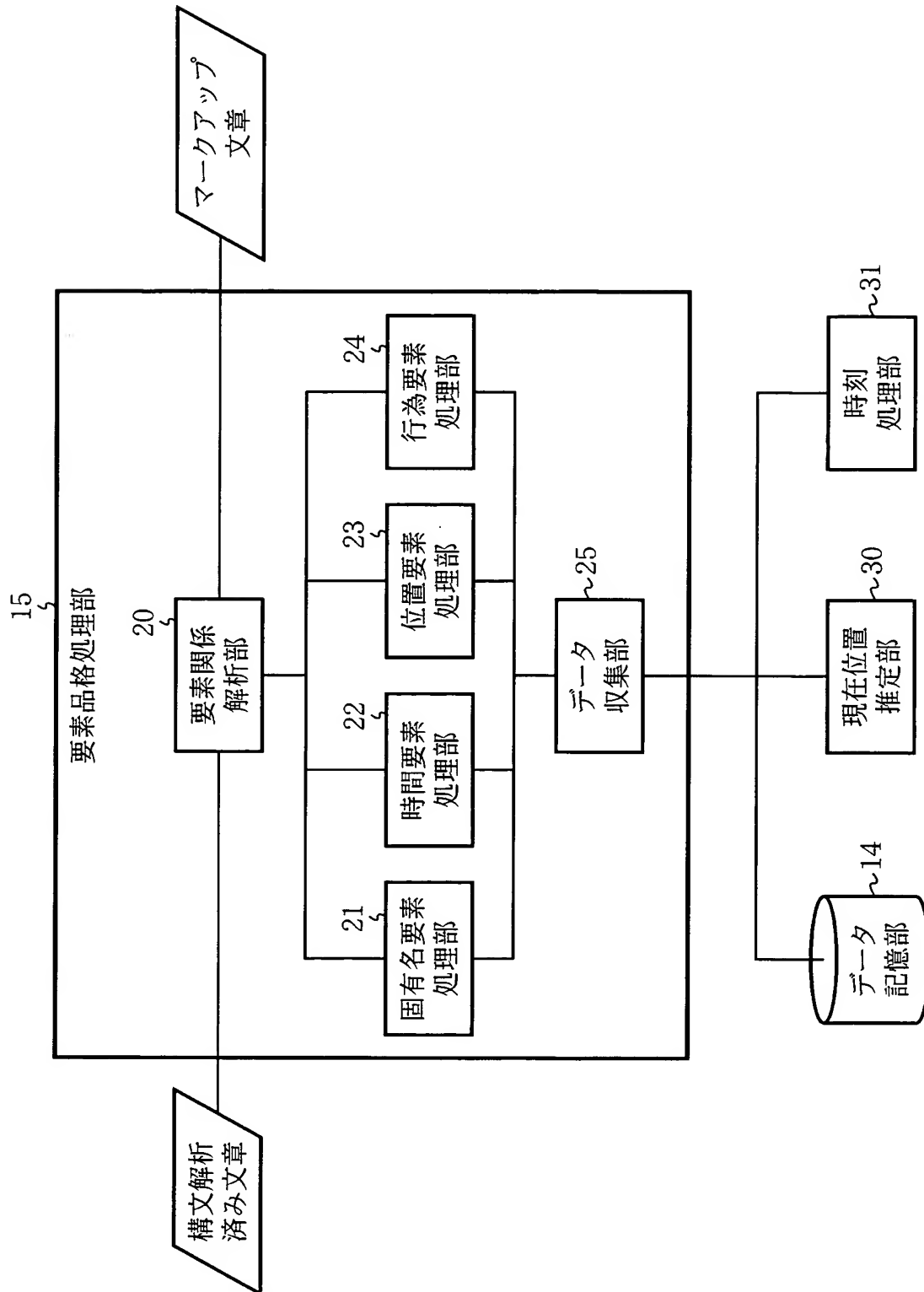
【図 4】



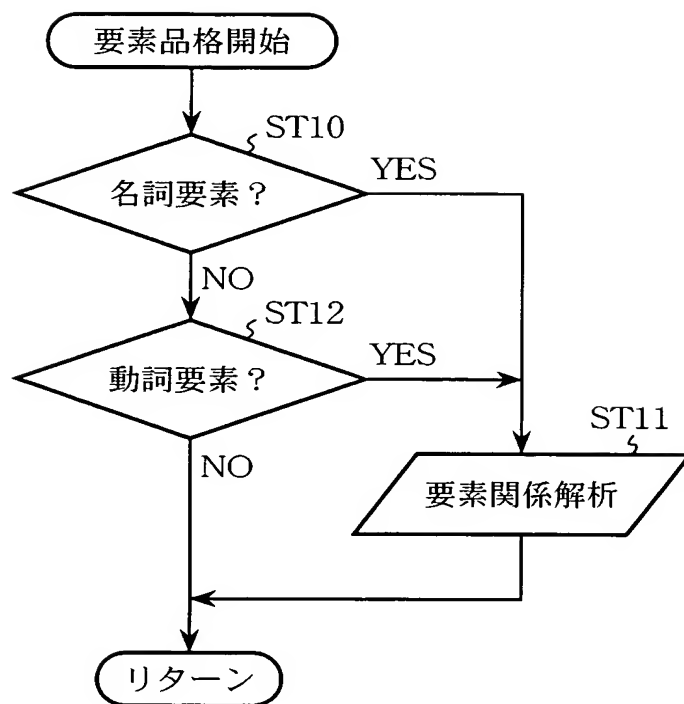
【図 5】



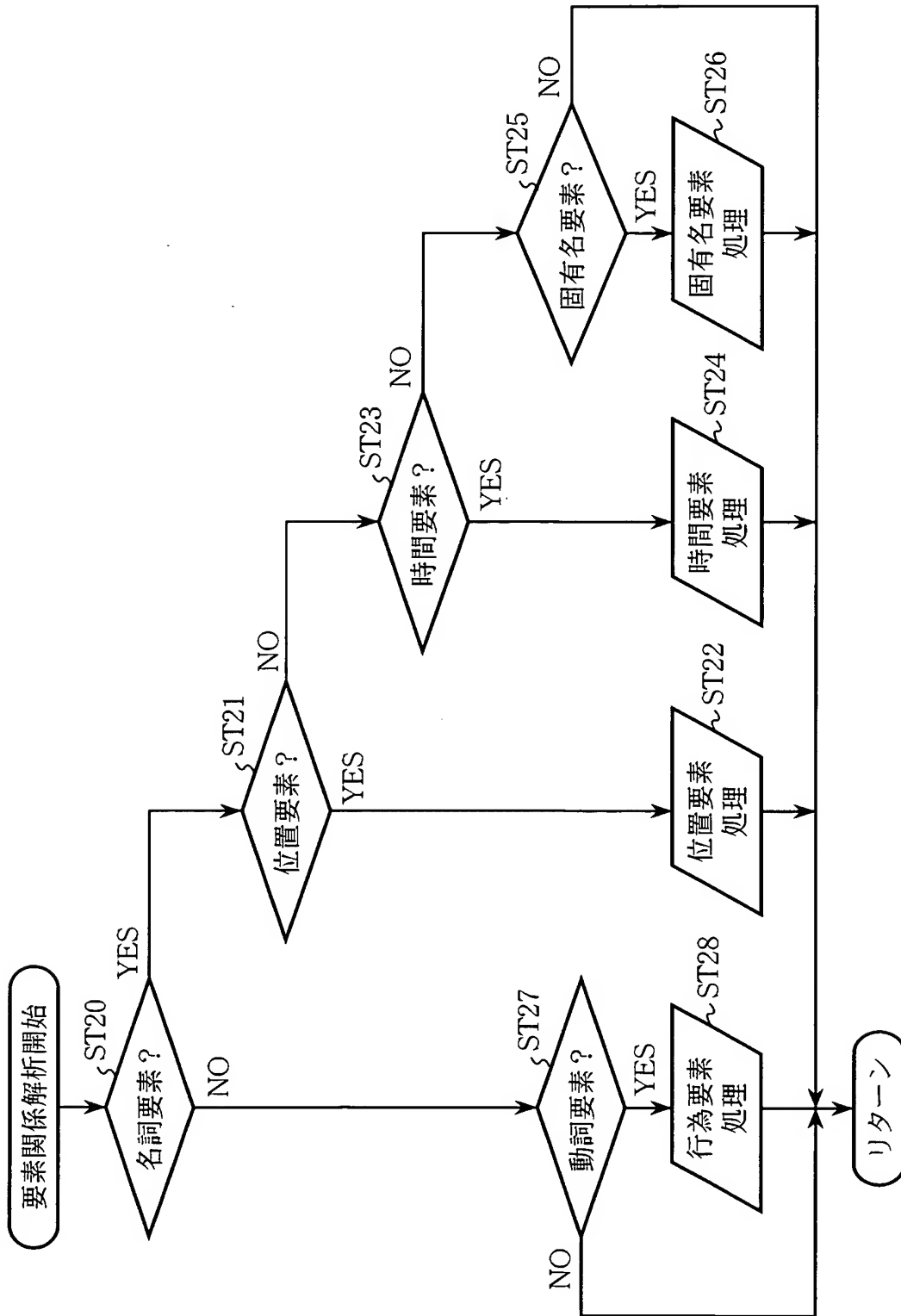
【図 6】



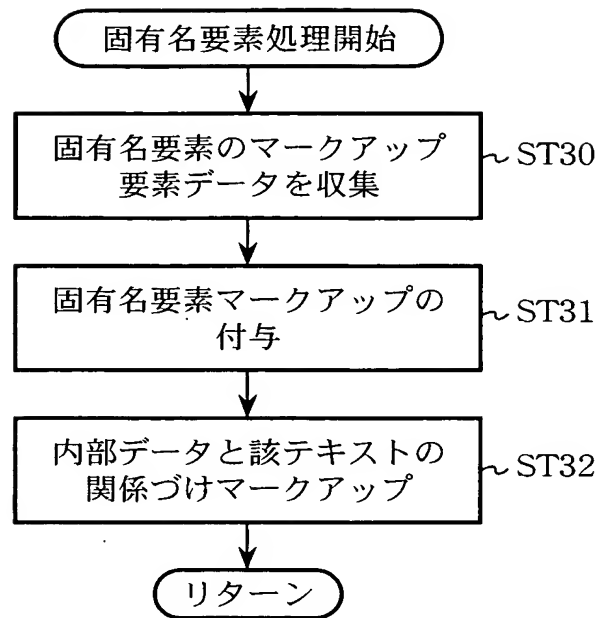
【図 7】



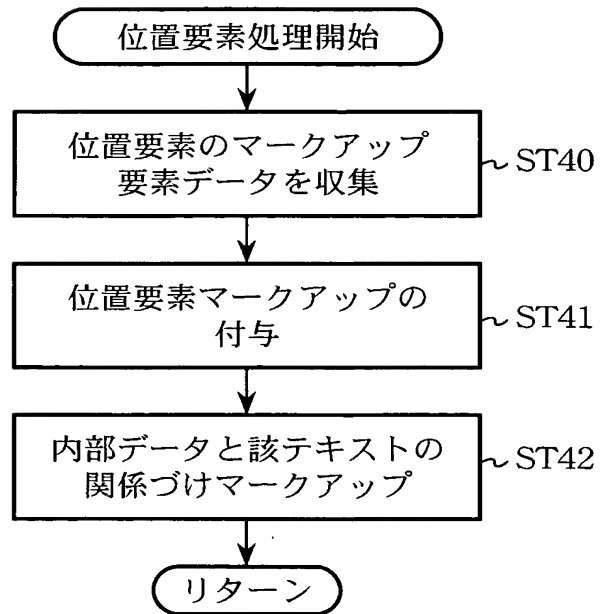
【図 8】



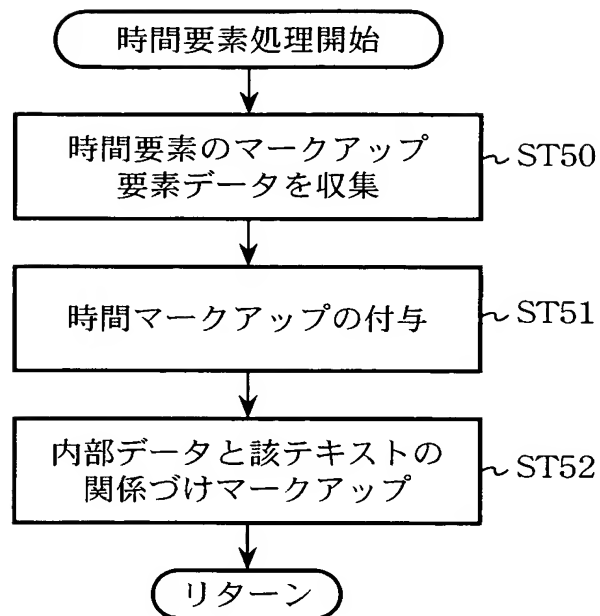
【図 9】



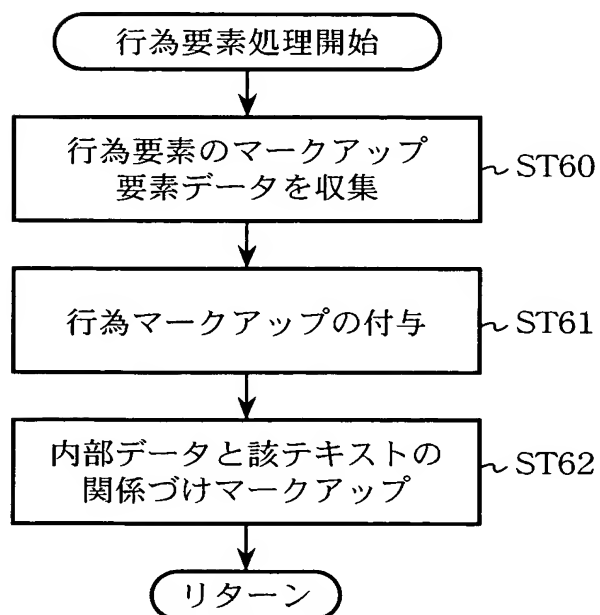
【図 10】



【図 11】



【図 12】



【書類名】 要約書

【要約】

【課題】 文章の意味及び内容を確実に伝えることのできる文章情報処理装置を提供する。

【解決手段】 プレーン文章入力部 10 から入力されたプレーンな文章を、辞書記憶部に記憶された辞書を用いて形態要素解析してトークンに分解する形態要素解析部 12 と、この形態要素解析部 12 で得られたトークンの品詞を解析して意味のある語を含む構造化された文章生成する構文解析部 13 と、この構文解析部 13 で生成された構造化された文章に含まれる意味のある語に対して、データ記憶部 14 に該語に関係付けて記憶されているデータを付加するマークアップを行う要素品格処理部 15 と、要素品格処理部 15 で生成されたマークアップ文章を出力するマークアップ文章出力部 17 とを備えている。

【選択図】 図 1

特願 2 0 0 3 - 1 0 6 8 7 8

出 願 人 履 歴 情 報

識別番号

[0 0 0 0 0 6 0 1 3]

1. 変更年月日

1 9 9 0 年 8 月 2 4 日

[変更理由]

新規登録

住 所

東京都千代田区丸の内 2 丁目 2 番 3 号

氏 名

三菱電機株式会社